A Position Paper on Extreme-scale Uncertainty Quantification Methodologies

Charles Tong
Lawrence Livermore National Laboratory
tong10@llnl.gov

Introduction

Complex model simulations with uncertainty quantification (UQ) generally require orders of magnitude higher computational cost than their deterministic counterparts. As such, UQ analysis can benefit greatly from advances in high performance computing. In fact, it has often been claimed that one major driver for extreme-scale computing is to enable complex simulations with UQ. The objective of this paper is to scrutinize this claim by analyzing the computational requirements of different UQ methodologies against some characteristics of extreme-scale architectures. Our position is that the validity of this claim depends on the type of UQ methodology employed together with the distinction between extreme-scale capacity and extreme-scale capability. In particular, we maintain that, without additional in-situ analyses that require tighter couplings (and thus communication of large shared data) between independent simulations, purely non-intrusive UQ methods offer a weak case for needing extreme-scale *capability* (not capacity). The goal of presenting this position is to stimulate new thinking in taking advantage of extreme-scale capabilities in developing new strategies for rigorous and efficient UQ.

Our Argument

Non-intrusive or sampling-based UQ generally involves a large number of independent simulations with different uncertain parameter values. Suppose 1000 independent simulations are needed for a UQ study, and each simulation requires petaflop capability to give acceptable turn-around time. If an exaflop computer is available (that is, all processing units are physically co-located), a straightforward simulation scenario is to divide the machine into 1000 partitions, each one of which is responsible for a single simulation. An equally feasible alternative is to run the simulations on 1000 petaflop computers that may reside at different geographical locations and are connected via the *internet*. Thus, it is fair to say that non-intrusive UQ requires extreme-scale capacity, but not so much extreme-scale (co-located with fast interconnection network) capability, due to the "embarrassing parallelism" in sampling-based UQ. This observation has at least two ramifications: (1) exascale UQ (non-intrusive) can be performed on today's petascale computers and thus the 20 MW power consumption in exascale challenge is less critical; and (2) this alternative offers the flexibility of not having to rely on exascale transistor technology (which will likely use near-threshold supply voltage) and thus helps to improve resiliency (assuming that the older petascale technology, which uses higher voltage, offers relatively lower soft error rates) as well as lessen the priority to address exascale issues such as lower memory per core and deep memory hierarchy.

Some Counter Arguments and Responses

This section considers several possible UQ strategies to circumvent the argument given in last section. These strategies attempt to strengthen the appropriateness of sampling-based UQ methods for extreme-scale computing by introducing tighter coupling into UQ simulations. The objective of this section is to assess the strengths and weaknesses of these strategies.

<u>Counter Argument #1</u>: Most petascale simulations generate large amount (giga- or tera-bytes) of output data. Since analysis of the UQ simulations requires all simulation outputs to be gathered at one place, it is pertinent that all simulations to be performed locally.

<u>Response #1</u>: Even though large amount of output data may be created from a complex simulation, most of the time the large data set is post-processed to extract a much smaller set of data metrics for UQ analysis. Thus, gathering these extracted outputs from the geographically scattered computers should not significantly impede the efficiency of UQ analysis.

<u>Counter Argument #2</u>: Newer UQ methodologies seek to enhance efficiency (for example, in creating response surfaces or surrogates) by sampling mostly at "important" regions in the

parameter space—an approach called "adaptive or importance" sampling. These adaptive strategies require monitoring simulation results continuously and thus they greatly benefit from the availability of extreme-scale computers.

Response #2: Most often only a small data set from each simulation is used to determine the next set of sample points. As in #1 above, gathering the data sets from all simulations for adaptive analysis should not create significant inefficiency even when the simulations are run on computers that are not physically co-located (The same can be said about parallel numerical optimization where each function evaluation is a petascale simulation.)

Counter Argument #3: Sampling-based UQ can be made more efficient by grouping them together using the concept of "single program multiple data" paradigm, that is, running one program on multiple instances corresponding to a number of sample points, so that some information (for example, physics lookup tables, Krylov vectors, preconditioners) may be shared among instances.

Response #3: This scenario is feasible, but may not be practical. For example, different instances may require different lookup tables. Also, the time-steps may be very different between instances that it may not be possible to share lookup tables, Krylov vectors, or preconditioners.

Counter Argument #4: Sampling-based UQ methods may benefit (need smaller sample size) from simulations that also compute derivatives with respect to the uncertain parameters (for example, using automatic differentiation), and these "loaded" simulations are more tightly coupled and more computationally intensive to require exascale capability.

Response #4: This scenario provides a stronger counter argument than the previous ones (actually this belongs to the class of 'semi-intrusive' methods). In practice, unless there are hundreds or thousands of uncertain parameters, each loaded simulation may not need exascale capability.

Counter Argument #5: Interval-based UQ methods that propagate epistemic uncertainties via overloaded data types (that encapsulate probability bounds instead of fixed values) and innovative interval arithmetic operations are much more computationally intensive and sufficiently coupled to justify exascale capability.

Response #5: It is unclear at this point how useful these methods are for complex models since current interval-based methods do not provide tight enough uncertainty bounds at the output. Nonetheless, further advances in this class of methods may make them more appealing.

Counter Argument #6: There are situations that a large data set is generated from each simulation and this data set can only be reduced by comparing it against the data sets from all other simulations via, for example, data mining techniques. This in-situ analysis requires that all simulations be co-located to avoid sending huge data sets across the communication network.

Response #6: This non-intrusive UQ analysis coupled with other in-situ analysis is by far the strongest case for needing extreme-scale computing, and thus should be explored further.

Further Discussions

The discussion above shows that sampling-based UQ methods are in general too decoupled to take advantage of extreme-scale capabilities. This section suggests three possible directions in UQ research and development that can better exploit (but not for the sake of) extreme-scale capabilities:

1. Integrate sampling-based UQ with other important tasks so that together they require tighter coupling between all simulations (e.g. in-situ analysis of large data sets across all simulations).
2. Intrusive UQ methods (e.g. based on generalized polynomial chaos: Xiu and Karniadakis, SIAM J. Sci. Comput. 24(2)) are natural candidates for extreme-scale computing. Major hurdles are the tedious tasks of re-formulation and re-implementation. Additional issues are resiliency and communication overhead. Advances in intrusive methods to address these hurdles and issues may make them more attractive on extreme-scale machines.
3. Mixed intrusive/non-intrusive UQ methods, that provide more flexibility in re-formulation and re-implementation and yet are tightly coupled, are good candidates for extreme-scale computers.